AIntelia

COVID-19 Face Mask Detection with Pre-trained Deep Learning Models

Salma Alraid[1*], Yasin Ortakci[2]

[1]Karabuk University, Karabuk, Turkey, salma110010@gmail.com, ORCID: 0000-0002-6817-3931*

[2]Karabuk University, Karabuk, Turkey, yasinortakci@karabuk.edu.tr, ORCID: 0000-0002-0683-2049

The global pandemic known as COVID-19 puts huge pressure on researchers to use technological solutions to provide further protection mechanisms. Face masks are one of the most important protection mechanisms among other health protocols. This study detects the wearing mask classification problem using four CNN models: VGG16, ResNet50V2, IncptionV3, and MobileNetV2 based on the Transfer Learning and also makes a fair comparison among their performance. The proposed models enhance the classification of wearing masks into three classes; without the mask, the correct wearing of the mask, and not the correct wearing of the mask. The four-transfer learning models of CNN architectures were used to train, test, and validate based on the image dataset. The results reveal that the proposed models have performed the classification task to detect the condition of wearing a mask. VGG16, ResNet50V2, and MobileNetV2 models achieved the same accuracy level of 99%, while the IncptionV3 achieved a little bit lower accuracy at 97%.

***Keywords**: Covid-19, Pre-Trained Models, Mask detection, Transfer learning.*

1. Introduction

Preventive measures in serious cases such as the spread of epidemics and infectious diseases are one of the most important means to be followed to reduce and control the situation. The process of wearing face masks is a low-cost and very effective precautionary measure in preventive cases, as we witnessed recently in the Covid-19 pandemic. Given the seriousness of the situation, governments and global and international health organizations have obliged people to wear masks. This process was accompanied by common mistakes in the process of wearing the mask, whether intentionally or unintentionally, and for the purpose of controlling the matter, researchers and engineers built smart models that have the ability to monitor and classify people on the basis of wearing a mask [1]. Assigning humans to the process of enforcing and following people who wear a mask has a very high cost, especially in a crowded public place. Fortunately, it is possible to employ technology as a replacement for humans. Around December 2019, a new coronavirus disease (COVID-19) was discovered in Wuhan, China (Hubei Province), and it has been spread to many other nations. Following the World Health Organization's (WHO) declaration of COVID-19 as a pandemic on 11 March, 2020, some nations implemented measures like limiting international travel [2]. The mask-wearing check system should be implemented right once to replace this new work with computerized, unsupervised work. It was anticipated that introducing this would lighten employees' workloads. The mask-wearing check system allows us to carry out mask-wearing activity at a lower cost while maintaining higher quality and accuracy. So, Artificial Intelligence (AI) can be defined as a computer system that can handle a large amount of data to perform tasks such as image classification, and decision-making [3]. In this digital age, AI plays a crucial role in the medical field due to its ability to cope with large clinical data [4]. Furthermore, deep learning is defined as part of a large group of methods of neural networks and is also known as representative science [5].

The rationalization of using preventive measures to reduce infection with the Coronavirus or any infectious diseases through the respiratory system is our first motivation in presenting this study. The unexpected COVID-19 outbreak has severely damaged both the health system and the global economy [6, 7] and put several countries in a vulnerable position. To lessen the impact of the epidemic, governments all around the world started to educate their citizens about the importance of health protocols. Warn the face mask is one of the most crucial health precautions. Despite the noticeable decline in the impact of the Coronavirus, we, as researchers, must continue to develop various ways to be ready in the event of a repeat of the 2019 scenario.

Many smart models can identify people who do not wear face masks, but the main problem is the people who do not wear the mask correctly, as they leave a part of the nose or mouth exposed. Wearing a mask this way is an unintended misleading of the devices, as they classify them as wearing masks. In this research, we aim to adapt models so that they can solve this issue.

In this study, we presented four deep learning pre-trained models of Inspectionv3, MobileNetV2, ResNetV2, and VGG-16. At the same time, we used a large number of images of people wearing the face mask incorrectly to train the models. Also, we fed models with the images of both wearing the mask correctly and not wearing it at all. The aim of this study is to record a higher performance among the four used models on the face mask classification problem.

2. Related Works

This section offers a non-exhaustive sample of works published on the face mask detection problem. We have selected several studies that dealt with the same deep learning models that we intend to use in this paper.

Sommana and Kitiyakara [8] presented a two-stage face mask detection problem. The first stage extracts the head region from the total body image by using the joint of PyramidKey and RetinaFace algorithms with ResNet50, MobileNetV11.0, MobileNetV20.25 and MobileNetV21.0 backbone models. The second stage classified the faces they wore the masks or not based on MobileNetV2. AIZOO and Moxa 3K were used as benchmarks datasets; both were labeled to two classes wearing and not wearing masks with a respectable number of images. The model achieved high results with AIZOO dataset while the results were extremely poor with Moxa 3K dataset because of unbalancing between these two classes. Then, they made a fair comparison with state-of-art methods results which demonstrated the high performance of the PyramidKey-MobileNetV11.0 model by achieving F1-score accuracies of 95.07 in the without mask class and 95.79 in the wearing mask class.

Farady et al. [9] proposed a model that combined both mask detection problems and head temperature using a deep learning detection approach. The project involves; reducing the spread of coronavirus infections, checking the people who enter public places for wearing masks, and having head temperature in real-time. Three modules were presented in this study: 1) object detection for detecting the 3 positions of the mask-wearing (with mask, without mask, or incorrect mask). 2) Head temperature detection with one class classifier to detect the head position. 3) Combine model gathering the two modules. The first two modules depended on One-stage object detection RetinaNet, this backbone network used ResNet50 as a convolutional neural network to extract the important features from input images. The results show a confidence score of 81.31% for real-time videos. Expanding the dataset by adding more training and a variety of thermal head or facial photos are suggestions for future work.

Han et al. [10] presented a single-shot detector (SSD) that is used for mask detection in a supermarket for the substantial risk of infection in these crowded spaces. The paper's contribution can be summarized as using SSD to improve the speed of real-time detection, proposing a Feature Enhancement Module (FEM) for smaller object detection, and construction of a large-scale image dataset. VGG-16 network was used with a lightweight backbone to ensure rapid detection of small objects.

Joshi et al. [11] provided a deep learning-based method for finding facial masks in videos. The suggested framework makes use of the Multi-Task Convolutional Neural Network face detection model to recognize faces and the facial landmarks that correspond to them that are visible in the video frame. A neoteric classifier uses the MobileNetV2 architecture as an object detector for finding masked regions to analyze these facial images and cues. The suggested framework was evaluated using a dataset made up of films documenting how individuals move about in public areas while adhering to COVID-19 safety standards. The proposed method achieved 86.6% accuracy and 87.8% recall.

Rahman et al. [12] suggested simulating a smart city and how to monitor and deal with people who do not wear masks in light of the Corona pandemic. The process begins with creating an integrated network of monitoring devices to monitor the movement of people. All captured videos are analyzed by systems connected to the network. This system is a smart device that has the ability to distinguish between people wearing face masks or not. The basic design of the smart models is based on a convolutional neural network, which consists of 17 layers, including 6 bypass layers while it was trained on an image dataset extracted from multiple websites. The result shows high accuracy, 98.7%, for distinguishing a mask-wearing person from a non-mask-wearing person.

Li [13] classified the wearing mask cases into three classes: not wearing a mask, wearing a mask, and not wearing a mask correctly. Two different datasets were used in this research, and the accuracy of the first test was 94.52% by using the InspectionV3 model while the second test achieved 96.68% after preprocessing operations.

3. Methodology

This study aims to employ the CNN pre-trained models: IncpectionV3, MobileNetV2, ResNet50, and VGG-16, for classifying the people into three groups wearing a mask, not wearing a mask, or wearing a mask incorrectly. As shown in Figure 1, in the first stage we load the images, then increase the image numbers with the augmentation method. The second stage is adjusting the weights of constructed models by the Transfer Learning (TL) layer. The criterion on which this paper is based is to know the highest prediction that can be achieved among the models used. The data was separated into 80/20 splitting rates for all used models for the training phase. During the training, models tuned their weights based on TL, while the evaluation step pushed the model's foreword to get more suitable weights. In the end, the classification step was executed to classify the classes.

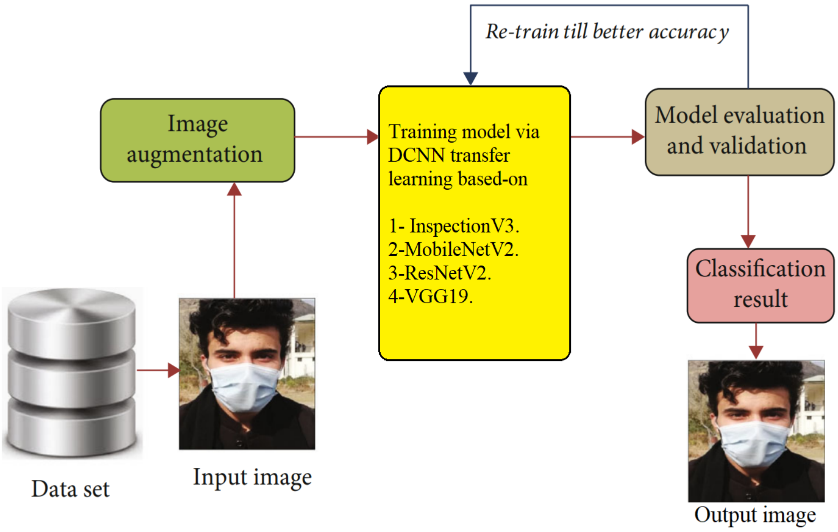

***Figure 1**. Illustrate the diagram of this study.*

A. Inception-V3:

The model is known without changing the structure, as the model consists of groups that each hold different numbers of convolution layers, MaxPooling layers, and dropout layers, fully connected layers, as well as the output pass through the SoftMax (Figure 2). We used TL InspectionV3, where the weights have already been trained using ImageNet dataset. The model structure consists of a cover of transferred learned InspectionV3, average pooling, a flattened layer, and two dense layers. The cover ran once with the first epoch to tune the weights of the total model then the head of the constructed model was placed on top [14]. The convolution layers were used for feature extraction from images the MaxPooling layers were used during the extraction operation to reduce the high demotions of extracted features. Also, drop out technique was used to prevent over fitting, where 5% of the network weights were dropped for each iteration. In the last stage, we used a dense layer with SoftMax regression equation and three output classes for classification issues.

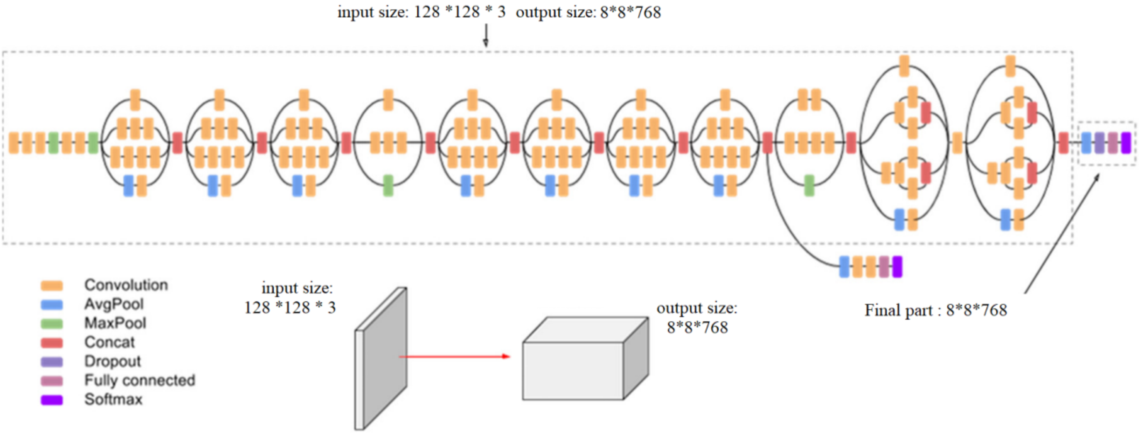

***Figure 2**. Explain the IncptionV3 design [14].*

B. MobileNetV2:

We used a MobileNetV2 pre-trained model by ImageNet dataset as a base network and then replaced the base network's top layer with a convolution layer and a SoftMax classifier. We applied dropout to reduce overfitting and used Adam Optimizer with a very low Learning Rate (LR), 0.0001. The pre-trained MobileNetV2 was used to extract features and tune the value of weights. SoftMax classifier was used to classify features. We trained this new model in two stages using Adam Optimizer of different LR. The constructed model layers with iterations are illustrated in Table 1.

***Table 1.** The constructed model of MobileNetV2.*

Iterations No.	Used layers
For the first iteration	MobileNetV2 network pre-trained by ImageNet
For all iterations	Conv2D
	AveragePooling2D
	Flatten
	Denes(128)
	Dropout (0.5)
	Denes(3)

C. ResNetV2:

As a base network, we used a ResNetV2 model that had been trained on the ImageNet dataset. The top layer of the base network was run once in the first iteration then it was replaced with the constructed model layers and a Softmax classifier. To lessen overfitting, we applied a dropout of 0.5 averages while we used Adam Optimizers with a small L.R., 0.0001. The entire framework is illustrated in Table 3.

Table 2. *The constructed model of ResNetV2.*

Iterations No.	**Used layers**
For the first iteration	ResNetV2 network pre-trained by ImageNet
For all iterations	Conv2D
	AveragePooling2D
	Flatten
	Denes(128)
	Dropout (0.5)
	Denes(3)

D. VGG19:

We used a VGG19 trained on the ImageNet dataset as our foundation network. A common convolution layer and a Softmax classifier were added to the base network's top layer to replace the original structure, as the model structure was listed in Table 3. We simultaneously added dropout to the recently introduced conv2d to reduce overfitting. The pre-trained VGG19 and the Softmax classifier were used for feature extraction and classification, while 0.0001 training rate was used.

Table 3. *The constructed model of VGG-16.*

Iterations No.	**Used layers**
For the first iteration	VGG-16 network pre-trained by ImageNet
For all iterations	Conv2D
	AveragePooling2D
	Flatten
	Denes(128)
	Dropout (0.5)
	Denes(3)

4. Experimental Study

A. VGG19:

The usage dataset was collected from Kaggle [15]. The final blend includes 8982 images divided into three subfolders: correct mask, incorrect mask, and without mask, each folder holds 2994 images of people that belong to such a labeled class. All images are in RGB colors and PNG format, while the images' sizes varied (higher or lower than (256Px*256Px)), as shown in Figure 3. The data were preprocessed before feeding it to the model: we reshaped the dimensions of the images to be (128Px*128Px*3) which reduced the images' sizes smoothly without affecting the features and lowered the computations and complexity. On the other hand, while 15% of the dataset remained original, 20% of the dataset was rotated, 15% was zoomed, 20% was shifted to the left, and 20% was shifted to the right.

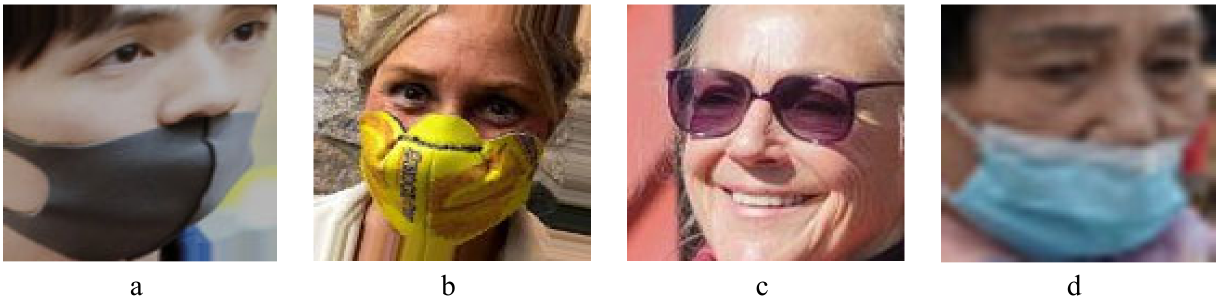

***Figure 3**. Sample Images from dataset usage, a: wearing mask incorrectly, b: wearing correctly, c: without mask, d: wearing incorrectly.*

B. Evaluation Metrics:

There are several metrics extracted through the confusion matrix to evaluate the machine learning models' achievement. In this study, we evaluate the presented models by the most common evaluation metrics: Accuracy, Recall, Precision and F1-score.

Accuracy is the most often used parameter for assessing classification performance. This measure computes the proportion of properly identified samples and is represented by Equation (1) [16].

$$accuracy = \frac{TN+TP}{T+F} \tag{1}$$

Where TP is the True Positive, TN is the True Negative, T is the total True (True Positive and True Negative together) and F total False (False Positive and False Negative together).

The Recall or sensitivity refers to the rate of true positive to the positive classes' samples that are appropriately labeled. Equation 2 shows the Recall mathematical representation:

$$recall = \frac{TP}{TP + FN} \tag{2}$$

Precision could be the opposite of Recall, where it deals with negative values i.e refer to the rate of the true negative to the negative classes samples that are appropriately labeled or how can the model detect the negative instances.

$$precision = \frac{TN}{TN + FP} \tag{3}$$

The last common measurement metric is F1-score which extracting from the values of Recall and Precision. Perfect score of this parameter means that the model leads the positive class; the equation of F1-score is shown below.

$$F1 - score = 2 * \frac{precision * recall}{precision + recall} \tag{4}$$

C. Experiment:

The training was done only for the classification model using the generated dataset, 30 epochs were carried out on 80% of the dataset as training samples and 20% as validation samples. Augmentation was preprocessed by randomly flipping, rotating, and zooming. We used the categorical-cross-entropy loss function was used during the model's compilation. All experiments were conducted using a high-performance P.C., Acer NITRO version core i5 which was adapted for running the codes with GPU computations by NVIDIA GEFORCE RTX 3060 with 6 GB and 16 GB memory.

D. Results:

Based on the selected metrics, the performance of InceptionV3, MobileNetV2, Resnet50v2 and VGG19 models has been calculated and listed in Table 5. We showed the accuracy and loss error of the four TL models for train and testing in Figure 4. We evaluated the success of models using accuracy, precision, Recall, and F1-score, as well as taking the execution time into account for each model. In Table 5, all models achieved extremely high accuracies, 0.99, except the Inspection model, which obtained 0.97. VGG-16 outperforms other models, which achieved 1.00 in each of the precision of the correct class, Recall and F1-score in the incorrect mask class. The ResNetV2 model got a good result compared with the MobileNetV2 and VGG-16 models but with exceptionally long execution time since its model construction has a huge complexity. We also detected that the weakest results were recorded in the precision column of the incorrect mask class. The reason is the difficulty of seeing the small visible parts, especially in the remote images captured (Figure 3.d). Table 4 shows the accuracy of the three models, MobileNet-V2, ResNet-V2, and VGG-16, which have achieved 99% outperforming InceptionV3. MobileNet-V2 has outperformed the other models regarding the run time achieving 544.7 sec., followed by the InceptionV3 with 559.32 sec. and 640.3 sec. for VGG-16, with the worst run time of 953.26 by the ResNet-V2. Hence, according to the accuracy and running time, the MobileNet-V2 outperforms the other models (Inception-V3, ResNet-V2, and VGG-16). Eventually, this section proves the performance evaluation of the models with reliable methods recommended.

Table 4. *Experiment results of the four deep learning models with TL, Acc. refers to (accuracy).*

Methods	**Calculation Time in second**	**Acc.**	**Correct-Mask**			**Incorrect-Mask**			**without-Mask**		
			Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
InceptionV3	559.32	0.97	0.97	0.91	0.94	0.93	0.97	0.98	0.97	0.99	0.98
MobileNetV2	544.7	0.99	0.99	0.98	0.98	0.98	0.99	0.99	1.00	0.99	0.99
ResNetV2	953.26	0.99	0.99	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.99
VGG-16	640.3	0.99	1.00	0.98	0.99	0.98	1.00	1.00	0.99	0.99	0.99

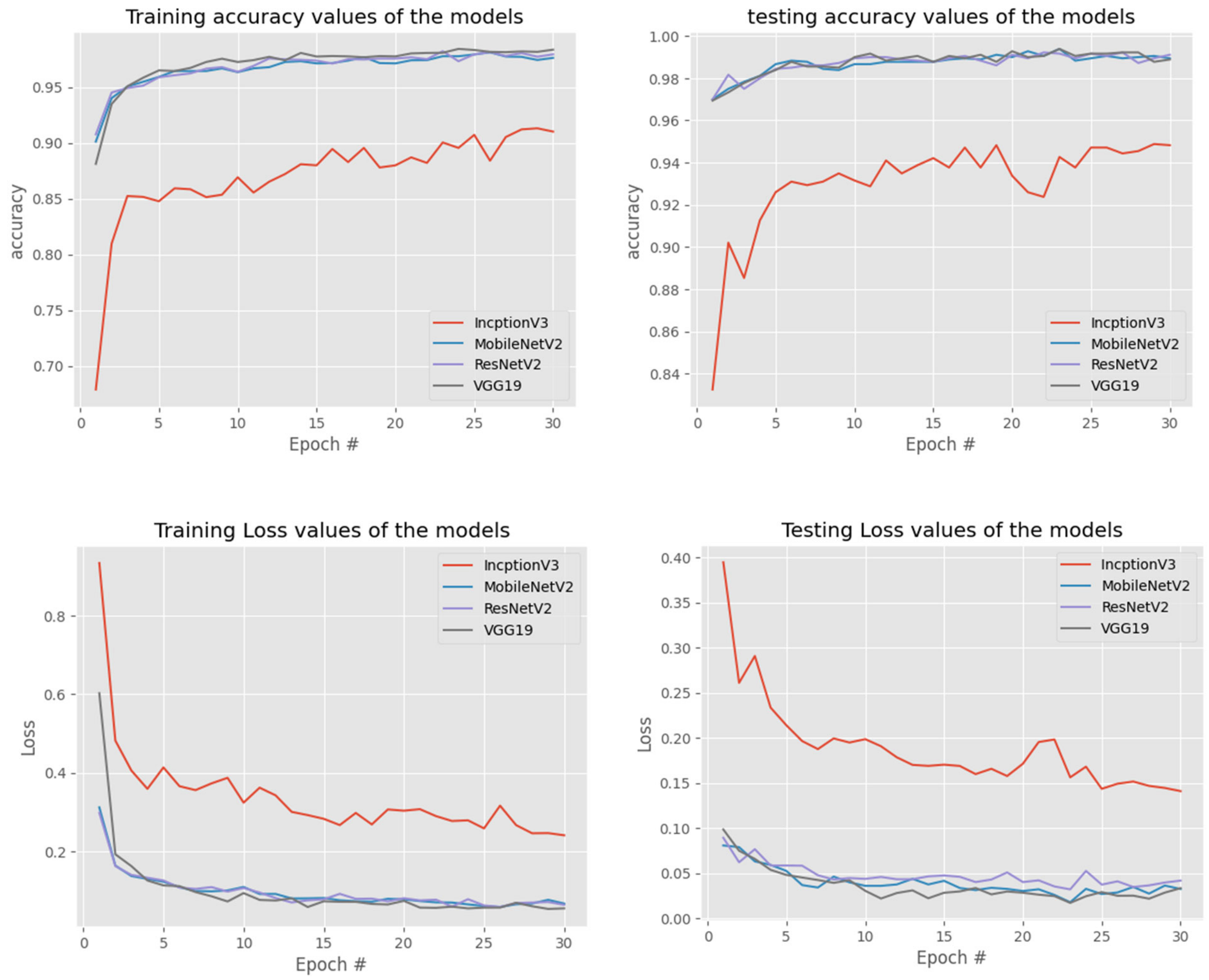

***Figure 4**. The loss and accuracy graphic of the four models.*

5. Conclusion

In this paper, we used the ResNet50V2, InceptionV3, VGG-19 and MobileNetV2 models as our baselines. We compared the classification performance accuracy and the running time beside to the recall, precision and f1-score on mask-wearing problems. According to Table 5, all the proposed TL models achieved high accuracy. We used the adaptive optimizer with the smallest LR value, 0.0001, while all models used 30 epochs and 32 batch sizes. The training was done only for the classification model using the generated dataset. 30 epochs were carried out on 80% of the dataset as training samples and 20% as validation samples. The models classified 8982 images from three classes: correct mask, incorrect mask and without a mask, each consisting of 2994 images. VGG-16 achieved the highest result, followed by MobileNetV2 and ResNet50V2, while InspectionV3 came in the last. The results in the three pre-trained models prove that the TL model can be used in mask detection problems. In the future, we suggest focusing on increasing the accuracy of the classification in relation to the incorrect wearing of the mask, by increasing the number of remote images of people wearing the mask incorrectly.

REFERENCES

[1] Kumar, A., Kalia, A., Sharma, A. and Kaushal, M., "A hybrid tiny YOLO v4-SPP module based improved face mask detection vision system". Journal of Ambient Intelligence and Humanized Computing, pp.1-14, Oct. 2021.

[2] W.H. Organization: Who director-general's opening remarks at the media briefng on COVID-19— 11 march 2020. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 (2020). Accessed: 8 December 2020.

[3] S. Yeasmin, “Benefits of Artificial Intelligence in Medicine," 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), pp. 1-6, Riyadh 2019.

[4] Aashay Pawar. “Medical Advantages of Artificial Intelligence", International Journal of Innovative Science and Research Technology (IJISRT), vol. 5 no. 1, pp. 599-604, 2020.

[5] I. Farady, C. -Y. Lin, A. Rojanasarit, K. Prompol and F. Akhyar, "Mask Classification and Head Temperature Detection Combined with Deep Learning Networks," 2020 2nd International Conference on Broadband Communications, Wireless Sensors and Powering (BCWSP), pp. 74-78, Yogyakarta 2020.

[6] H. A. Rothan and S. N. Byrareddy, "The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak," J Autoimmun, vol. 109, no., p. 102433, Feb. 2020.

[7] T. Greenhalgh, M. B. Schmid, T. Czypionka, D. Bassler, and L. Gruer, “Face masks for the public during the covid-19 crisis,” The BMJ, vol. 369, Apr. 2020.

[8] B. Sommana et al., "Development of a face mask detection pipeline for mask-wearing monitoring in the era of the COVID-19 pandemic: A modular approach," 2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 1-6, Bangkok 2022.

[9] I. Farady, C. -Y. Lin, A. Rojanasarit, K. Prompol and F. Akhyar, "Mask Classification and Head Temperature Detection Combined with Deep Learning Networks," 2020 2nd International Conference on Broadband Communications, Wireless Sensors and Powering (BCWSP, pp. 74-78), Yogyakarta 2020.

[10] W. Han, Z. Huang, A. Kuerban, M. Yan, and H. Fu, "A Mask Detection Method for Shoppers under the Threat of COVID-19 Coronavirus," 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL), pp. 442–447, Chongqing 2020.

[11] A. S. Joshi, S. S. Joshi, G. Kanahasabai, R. Kapil and S. Gupta, "Deep Learning Framework to Detect Face Masks from Video Footage," 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), pp. 435-440, Bhimtal 2020.

[12] M. M. Rahman, M. M. H. Manik, M. M. Islam, S. Mahmud and J. -H. Kim, "An Automated System to Limit COVID-19 Using Facial Mask Detection in Smart City Network," 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), pp. 1-5, Vancouver 2020.

[13] Li, Y., “Facemask detection using inception V3 model and effect on accuracy of data preprocessing methods”. Journal of Physics: Conference Series, vol. 2010, no. 1, p. 012052, September, 2021.

[14] Patel, Khush. (2020) "Architecture comparison of AlexNet, VGGNet, ResNet, Inception, DenseNet" https://towardsdatascience.com/architecture-comparison-of-alexnet-vggnet-resnet-inceptiondensenet-beb8b116866d

[15] https://www.kaggle.com/datasets/vijaykumar1799/face-mask-detection

[16] M. Rahimzadeh and A. Attar, “A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2”, Informatics Med. Unlocked, vol. 19, p. 100360, 2020.